

CS Seminar

Database Scalability, Elasticity, and Autonomy in the Cloud

Prof Divy Agrawal
Department of Computer Science
University of California at Santa Barbara

Date:**December 16, 2011****Friday****2:00 pm****Venue:****Room 308****Chow Yei Ching Building****The University of Hong Kong****Abstract:**

Over the past two decades, database and systems researchers have made significant advances in the development of algorithms and techniques to provide data management solutions that carefully balance the three major requirements when dealing with critical data: high availability, reliability, and data consistency. However, over the past few years the data requirements, in terms of data availability and system scalability, from Internet scale enterprises that provide services and cater to millions of users has been unprecedented. Cloud computing has emerged as an extremely successful paradigm for deploying Internet and Web-based applications. Scalability, elasticity, pay-per-use pricing, and autonomic control of large-scale operations are the major reasons for success and widespread adoption of cloud infrastructures. Current proposed solutions to scalable data management, driven primarily by prevalent application requirements, significantly downplay the data consistency requirements and instead focus on high scalability and resource elasticity to support data-rich applications for millions to tens of millions of users. In particular, the "newer" data management systems limit consistent access only at the granularity of single objects, rows, or keys, thereby significantly trading-off consistency in order to achieve very high scalability and availability. But the growing popularity of "cloud computing", the resulting shift of a large number of Internet applications to the cloud, and the quest towards providing data management services in the cloud, has opened up the challenge for designing data management systems that provide consistency guarantees at a granularity which goes beyond single rows and keys. In this talk, we analyze the design choices that allowed modern scalable data management systems to achieve orders of magnitude higher levels of scalability compared to traditional databases. With this understanding, we highlight some design principles for data management systems that can be used to augment existing databases with new cloud features such as scalability, elasticity, and autonomy. In this talk we present recent advances that have been made to strike a middle-ground between the two radically different data management architectures: traditional database management systems where the data is treated as a "whole" versus modern key-value stores where data is treated as a collection of independent "granules".

About the Speaker:

Dr. Divyakant Agrawal is a Professor of Computer Science at the University of California at Santa Barbara. His research expertise is in the areas of database systems, distributed computing, data warehousing, and large-scale information systems. From January 2006 through December 2007, Dr. Agrawal served as VP of Data Solutions and Advertising Systems at the Internet Search Company ASK.com. Dr. Agrawal has also served as a Visiting Senior Research Scientist at the NEC Laboratories of America in Cupertino, CA from 1997 to 2009. During his professional career, Dr. Agrawal has served on numerous Program Committees of International Conferences, Symposia, and Workshops and served as an editor of the journal of Distributed and Parallel Databases (1993-2008), the VLDB journal (2003-2008) and currently serves on the editorial boards of the Proceedings of the VLDB and ACM Transactions on Database Systems. He has recently been elected to the Board of Trustees of the VLDB Endowment. Also, he served as the Program Chair of the 2010 ACM International Conference on Management of Data and served as the General Chair of the 2010 ACM SIGSPATIAL Conference on Advances in Geographical Information Systems. In 2011, Dr. Agrawal is organizing the NSF Workshop on the Science of Cloud Computing and will be serving as the General Co-Chair of ACM SIGSPATIAL Conference on Advances in GIS (ACM GIS'2011). Dr. Agrawal's research philosophy is to develop data management solutions that are theoretically sound and are relevant in practice. He has published approximately 300 research manuscripts in prestigious forums (journals, conferences, symposia, and workshops) on wide range of topics related to data management and distributed systems and has advised more than 30 Doctoral students during his academic career. Recently, Dr. Agrawal has been recognized as an Association of Computing Machinery (ACM) Distinguished Scientist and will be inducted as an ACM Fellow in June'2012. His current interests are in the area of scalable data management and data analysis in Cloud Computing environments, security and privacy of data in the cloud, and scalable analytics over social networks data and social media.

All are welcome!**For enquiries, please call 2859 2180 or email enquiry@cs.hku.hk****Department of Computer Science
The University of Hong Kong**